Vestiges of a DNA methylation system in *Drosophila melanogaster*?

The insect *Drosophila melanogaster* belongs to an atypical group of animals with no detectable genomic 5-methylcytosine[1–3]. We found, unexpectedly, that the *Drosophila* genome potentially encodes two proteins that resemble a cytosine DNA methyltransferase and a mammalian methyl-CpG-binding-domain (MBD) protein, respectively. The hypothetical DNA methyltransferase, dDNMT, is closely related to the pmt1/DNMT2 family identified in fission yeast[4] and mouse[5,6] (Fig. 1*a*). Attempts to transfer methyl groups from S-adenosylmethionine to DNA have not been successful using native members of this family, and bacterially expressed dDNMT was catalytically inactive *in vitro* (data not shown). Although the *dDNMT* transcript is present in embryonic, larval and adult flies (data not shown), the functional significance of the protein is unknown. Our report focuses on the *Drosophila* protein dMBD2/3, which is related to MBD2 and MBD3 of mammals[7] (Fig. 1*b*). We detected two transcripts encoding dMBD2/3 in *Drosophila* cultured-cell RNA, 0-hour to 4-hour embryonic stages and adults, suggesting widespread expression of the gene (data not shown). The shorter splice variant lacks part of the methyl-CpG-binding-domain homology (dMBD2/3Δ; Fig. 2*f*; ref. 8, and data not shown).

The carboxy-terminal amino acid sequence of dMBD2/3 resembles that of mammalian MBD2 and MBD3 (Fig. 1*b*), both of which belong to histone-deacetylase-containing corepressor complexes. MBD3 is within the Mi-2/NuRD complex[8,9], whereas MBD2 is within the MeCP1 complex[10]. Does dMBD2/3 also associate with *Drosophila* histone deacetylases (dHDACs)? We found that *in-vitro*-translated dHDAC1 (Fig. 2*a*) and endogenous dHDAC1 from a nuclear extract of *Drosophila* cells (Fig. 2*b*) were bound by immobilized dMBD2/3–GST fusion protein. An *in vivo* interaction was demonstrated by transiently co-expressing haemaglutinin (HA) epitope-tagged dMBD2/3 and dHDAC1 in mammalian cells (Fig. 2*c*); immunoprecipitates of HA-dMBD2/3 also contained dHDAC1. In addition, dMBD2/3 associated with the corepressor protein dMi-2 dMBD2/3 *in vitro* (Fig. 2*a*) and *in vivo* (Fig. 2*d*). Therefore, like mammalian MBD3, dMBD2/3 can interact with components of the Mi-2/NuRD histone deacetylase complex[11].

The methyl-CpG-binding domain of dMBD2/3 differs at sites that are highly conserved among vertebrate family members, including a nine amino acid deletion and an opa-like[12] repeat insertion. Is this because the methyl-CpG-binding domain is superfluous in an organism without CpG methylation? Or does it reflect structural or functional divergence in the evolutionary interval that separates insects and mammals? To distinguish between these alternatives, we sought an insect that has methylated DNA to ask whether its MBD2/3 orthologue resembled mammalian or *Drosophila* MBD proteins. The presence of 5-methylcytosine in the genome of the cricket *Acheta domesticus* was established by probing blots of *Hpa*II-digested single-stranded DNA with an antibody against the modified base[1,13] (Fig. 2*e*). An *A. domesticus* cDNA encoding an MBD2/3-like protein (aMBD2/3) was isolated by degenerate RT–PCR. The methyl-CpG-binding domain of aMBD2/3 (Fig. 1*b*) closely resembled the mammalian MBD2 and MBD3 domains, but differed from the

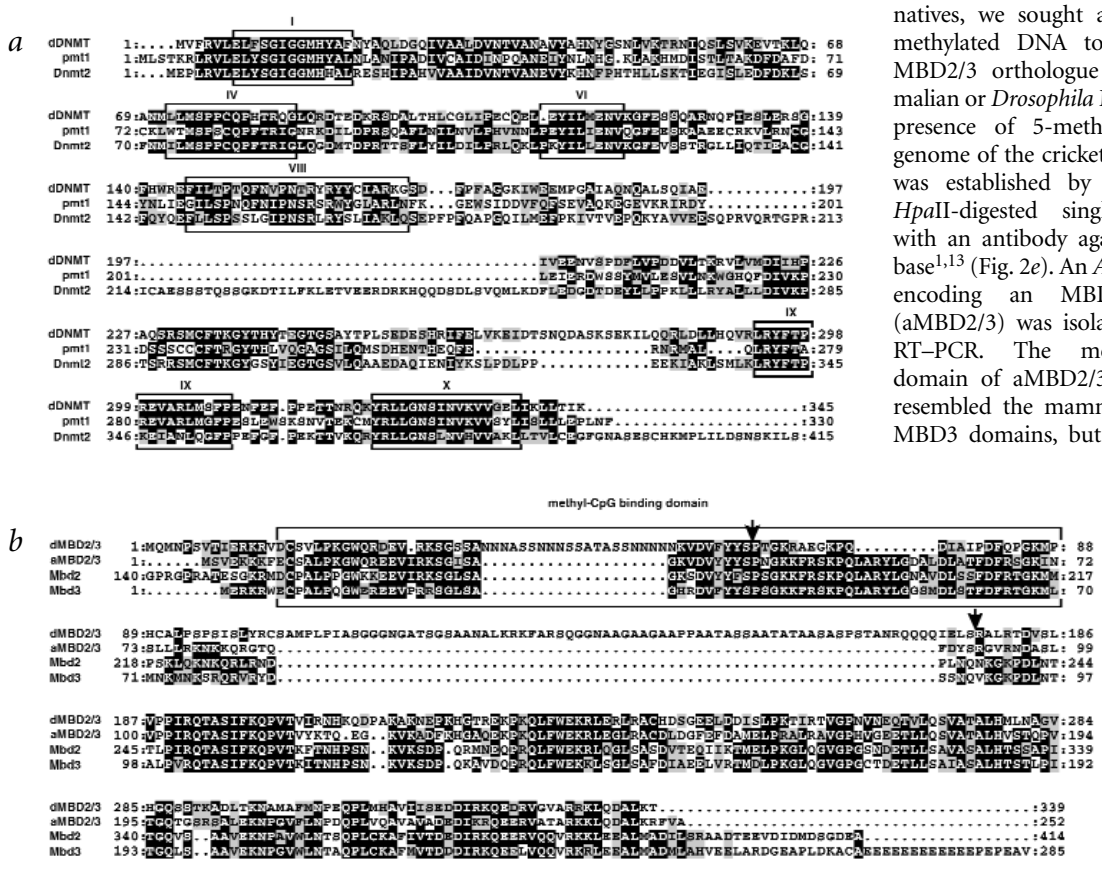

Fig. 1 Identification of candidate cytosine DNA methyltransferase and methyl-CpG-binding proteins in *Drosophila*. ***a***, Alignments of dDNMT with Pmt1 of *Schizosaccharomyces pombe*[4] and Dnmt2 of mouse[5,6]. Conserved cytosine methyltransferase domains I, IV, VI, VIII, IX and X (ref. 14) are boxed. ***b***, Alignment of dMBD2/3 and aMBD2/3 with Mbd2 and Mbd3 of mouse[7]. dMBD2/3 (LD4502; accession no. AA201537) was identified by BLAST search of EST databases with the methyl-CpG-binding domain. A fragment of aMBD2/3 was amplified by nested RT–PCR using degenerate primers, and the full cDNA sequence was completed by 5′- and 3′-RACE. The arrowheads indicate conserved intron positions in the corresponding genomic sequence of all species. Amino acids between the arrows are absent in dMBD2/3Δ and aMBD2/3Δ.

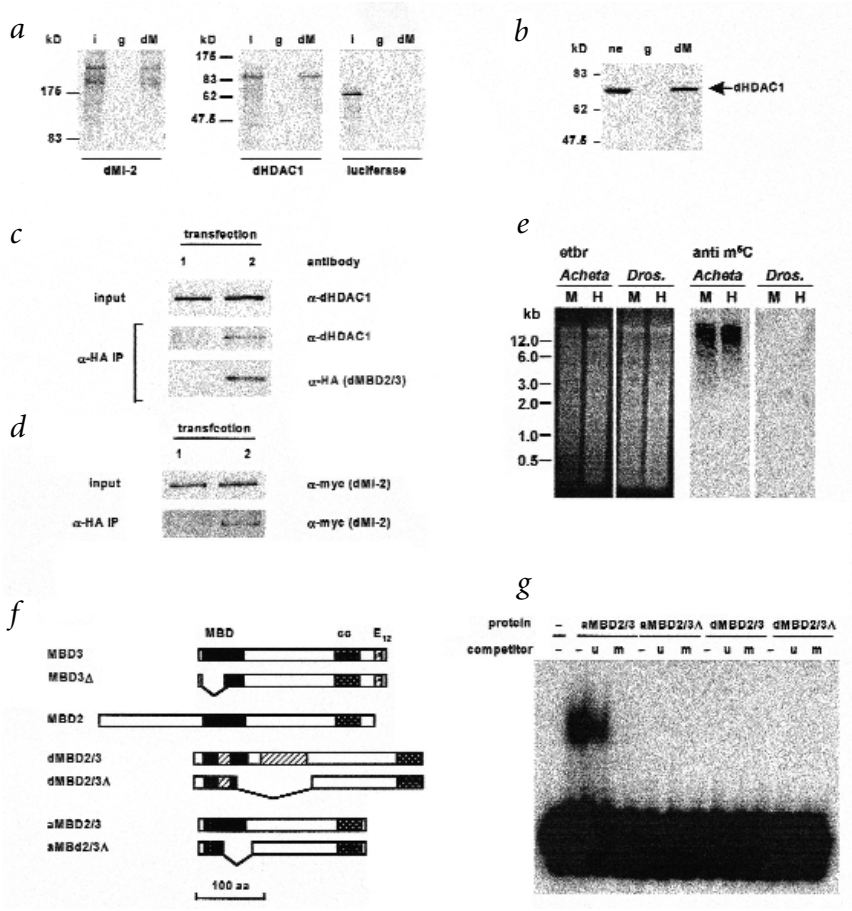

Fig. 2 Protein and DNA associations of dMBD2/3 and aMBD2/3. ***a***, Pull down[15] of *in-vitro*-translated dHDAC1 and dMi-2 by immobilized GST–dMBD2/3 (dM), but not by GST alone (g). Control luciferase is not bound by either protein. Input (i) shows 10% translation products labelled with ^{35}S-methionine before binding. The aMBD2/3 protein could also pull down dHDAC1 and dMi-2 (data not shown). The lower band in the dMi-2 lanes is a truncated dMi-2 product. ***b***, Pull down of endogenous dHDAC1 from *Drosophila* SL2 cell nuclear extracts by immobilized GST–dMBD2/3. dHDAC1 was detected on a western blot by anti-dHDAC1 antibody (A.L.B. and B.M.T., unpublished data). ***c***, Co-immunoprecipitation of transiently co-expressed dMBD2/3 and dHDAC1. Mouse L cells were transfected in parallel with 2 μg pMyc-dHDAC1 expression construct plus 4 μg vector pHM6 (transfection 1) or pHA-dMBD2/3 (transfection 2). Total extracted protein (input) or anti-HA immunoprecipitates (αHA IP) were probed on western blots with anti-HDAC (α-HDAC1) or anti-HA (α-HA) antibodies to detect exogenous dMBD2/3. HA-dMBD2/3 was immunoprecipitated using 3F10 anti-HA antibody (2 μg). Co-immunoprecipitated proteins were detected by anti-dHDAC antibody . ***d***, Co-immunoprecipitation of transiently co-expressed pHA-dMBD2/3 and pMyc-dMi-2. Vector alone (transfection 1) or pHA-dMBD2/3 (transfection 2) were transiently co-expressed with 2 μg pMyc-dMi-2. Western blots of total protein (input) or anti-HA immunoprecipitates were probed with 9E10 anti-myc antibody to detect dMi2. ***e***, Immunoblot of genomic DNA from *A. domesticus* and *Drosophila* after *Hpa*II or *Msp*I digestion[1]. Ethidium-bromide-stained agarose gels (left) were blotted and probed with a monoclonal anti-5-methylcytosine antibody[1,13] (right). ***f***, Variant forms of aMBD2/3 and dMBD2/3 proteins aligned with Mbd2 and Mbd3 of mouse. The methyl-CpG-binding domain (MBD), putative coiled-coil domain (cc) and a run of 12 glutamic acid residues (E_{12}) are indicated. Diagonally shaded regions of dMBD2/3 are not found in the other proteins. ***g***, aMBD2/3 (fused to GST) can bind[16] specifically to methylated DNA, but aMBD2/3Δ, dMBD2/3 and dMBD2/3Δ fusions (*f*) do not. The ^{32}P-labelled duplex probe was 5′–GATCG-GCGMGMGMGMGCC–3′. The unlabelled competitor was 250-ng methylated (m) or unmethylated (u) pCG11 (ref. 16).

dMBD2/3 domain (Fig. 1*b*), suggesting that the common insect ancestor of *Drosophila* and *A. domesticus* encoded an aMBD2/3-like protein. RT–PCR from *A. domesticus* RNA also identified a second transcript (aMBD2/3Δ) that is alternately spliced at the same positions as dMBD2/3 (Fig. 2*f*, and data not shown). Both insect MBD proteins, together with their major splice-variant forms (Fig. 2*f*), were tested for DNA binding. Only aMBD2/3 formed a strong complex with the methylated probe, and its binding was specifically competed by unlabelled methylated DNA (Fig. 2*g*). None of the other proteins, including dMBD2/3, bound to methylated DNA. Therefore, methyl-CpG binding by insect MBD2/3 proteins correlates with genomic DNA methylation.

Vertebrates produce proteins of the MBD2/3 family that either do or do not bind methylated DNA. Mouse Mbd2 (refs 7,10) and *Xenopus laevis* MBD3 (ref. 8) bind efficiently to methylated DNA, and mammalian MBD3 shows weak specific binding *in vitro*[7,8]. Both mammals and *X. laevis* produce an abundant splice variant of MBD3 that can no longer bind DNA (Fig. 2*f*; refs 7–9). We propose that the gene encoding aMBD2/3 resembles an ancestral member of the MBD2/3 family that encodes two corepressor functions. This is achieved by alternative splicing of a single transcript to yield DNA-binding and non-DNA-binding forms of the protein (Fig. 2*f*,*g*). As *Drosophila* appears to have lost CpG methylation, mutations that inactivate the methyl-CpG-binding domain of dMBD2/3 may have been allowed to accumulate. Whether divergence of the *Drosophila* domain has also given rise to novel functions is not known. The C-terminal domain of dMBD2/3, however, still associates with HDACs, presumably in the *Drosophila* Mi-2/NuRD complex. Persistence of the gene encoding dMBD2/3 without a functional methyl-CpG-binding domain is likely to reflect a requirement for its conserved C-terminal domain in corepressors that do not directly bind DNA.

Acknowledgements

We thank J. Müller for the dMi-2 plasmid; V. French for crickets; D. Finnegan for fly DNA; D. Macleod for comments on the manuscript; and J. Charlton, V. Clark, A. Greig and J. Davidson for technical assistance. This work was supported by grants to A.B. and B.M.T. from the Wellcome Trust. H.-H.N. is a Darwin Trust Scholar.

Susan Tweedie[1], Huck-Hui Ng[1], Andrew L. Barlow[2], Bryan M. Turner[2], Brian Hendrich[1] & Adrian Bird[1]

[1]Institute for Cell and Molecular Biology, University of Edinburgh, King's Buildings, Edinburgh, UK. [2]Department of Anatomy, The Medical School, University of Birmingham, Birmingham, UK. Correspondence should be addressed to A.B. (e-mail: a.bird@ed.ac.uk).

1. Tweedie, S., Charlton, J., Clark, V. & Bird, A. *Mol. Cell. Biol.* **17**, 1469–1475 (1997).
2. Urieli-Schoval, S., Gruenbaum, Y., Sedat, J. & Razin, A. *FEBS Lett.* **146**, 148–151 (1982).
3. Bird, A.P. & Taggart, M.H. *Nucleic Acids Res.* **8**, 1485–1497 (1980).
4. Wilkinson, C.R.M., Bartlett, R., Nurse, P. & Bird, A.P. *Nucleic Acids Res.* **23**, 203–210 (1995).
5. Yoder, J.A. & Bestor, T.H. *Hum. Mol. Genet.* **7**, 279–284 (1998).
6. Okano, M., Xie, S. & Li, E. *Nucleic Acids Res.* **26**, 2536–2540 (1998).
7. Hendrich, B. & Bird, A. *Mol. Cell. Biol.* **18**, 6538–6547 (1998).
8. Wade, P.A. *et al. Nature Genet.* **23**, 62–66 (1999).
9. Zhang, Y. *et al. Genes Dev.* **13**, 1924–1935 (1999).
10. Ng, H.-H. *et al. Nature Genet.* **23**, 58–61 (1999).
11. Kehle, J. *et al. Science* **282**, 1897–1899 (1998).
12. Wharton, K.A., Yedvobnick, B., Finnerty, V.G. & Artavanis-Tsakonas, S. *Cell* **40**, 55–62 (1985).
13. Barbin, A. *et al. Hum. Genet.* **94**, 684–692 (1994).
14. Kumar, S. *et al. Nucleic Acids Res.* **22**, 1–10 (1994).
15. Nan, X. *et al. Nature* **393**, 386–389 (1998).
16. Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L. & Bird, A.P. *Cell* **58**, 499–507 (1989).